

Received: 12 October 2023
Revised: 18 December 2023
Accepted: 30 December 2023
Published: 31 December 2023

Klasterisasi Desa di Provinsi Jawa Barat Berdasarkan Indeks Pembangunan Desa (IPD) Tahun 2021 Menggunakan Algoritma *K-Prototypes*

Siti Hasanah^{1, a)}, Irsyifa Mayzela Afnan^{1, b)}, Anwar Fitrianto^{1, c)},
Erfiani^{2, d)}, Alfa Nugraha^{1, e)}

¹*Program Studi Statistika dan Sains Data, Fakultas Matematika dan Ilmu Pengetahuan Alam, IPB University, Jl. Raya Dramaga, Kabupaten Bogor, Jawa Barat, 16680*

Email: ^{a)}stathasanah@apps.ipb.ac.id, ^{b)}irsyifamayzela@apps.ipb.ac.id, ^{c)}anwarstat@gmail.com,
^{d)}erfiani@apps.ipb.ac.id, ^{e)}alfanugraha@apps.ipb.ac.id

Abstract

Cluster analysis is a method used to group data with similar characteristics. There are various clustering methods adapted to different types of data. *K-Prototypes* is a clustering method that can be applied to mixed numerical and categorical data. The data used in this study are mixed numerical and categorical data derived from the Village Potential data in 2021. The aim of this research is to group villages in West Java based on variables from the Indeks Pembangunan Desa (IPD). Clustering using three clusters adapted to village status according to IPD resulted in 931 villages in cluster-1, 1880 villages in cluster-2, and 2104 villages in cluster-3. The characteristics of cluster-1 villages are villages that have adequate health and education facilities and good infrastructure conditions. Cluster-2 has an average numeric variable lower than cluster-1 but higher than cluster-3.

Keywords: Cluster Analysis, Mixed Data, West Java Villages, IPD, Village Development Index, *K-Prototypes*

Abstrak

Analisis kluster merupakan metode untuk mengelompokkan data dengan karakteristik yang mirip. Terdapat berbagai metode klasterisasi yang disesuaikan dengan tipe data. *K-prototypes* merupakan metode klasterisasi yang dapat diaplikasikan pada tipe data campuran numerik dan kategorik. Data yang digunakan pada penelitian ini merupakan data campuran numerik dan kategorik yang berasal dari data Potensi Desa Tahun 2021. Penelitian ini bertujuan untuk mengelompokkan desa-desa di Jawa Barat berdasarkan peubah dari Indeks Pembangunan Desa (IPD). Klasterisasi menggunakan tiga klaster yang disesuaikan dengan status desa menurut IPD menghasilkan 931 desa klaster-1, 1880 desa klaster-2, dan 2104 desa klaster-3. Karakteristik desa klaster-1 adalah desa-desa yang memiliki sarana kesehatan dan pendidikan yang jumlahnya memadai dan kondisi infrastruktur yang baik. Klaster-2 memiliki rerata peubah numerik di bawah klaster-1 namun diatas klaster-3.

Kata-kata kunci: Analisis Klaster, Data Campuran, Desa Jawa Barat, IPD, *K-Prototypes*

PENDAHULUAN

Analisis kluster merupakan salah satu metode analisis peubah ganda yang mengelompokkan objek berdasarkan kemiripan karakteristik yang dimiliki (Trianasari, dkk., 2021). Analisis kluster menurut strukturnya dibagi menjadi dua, yaitu hierarki dan non hierarki. Metode berhierarki digunakan ketika banyaknya kluster yang akan dibentuk tidak diketahui sebelumnya dan kurang efisien untuk data berukuran sangat besar. Sedangkan metode non hierarki digunakan jika kluster yang akan dibentuk sudah diketahui sebelumnya. Algoritma *k-prototypes* dapat digunakan untuk melakukan klusterisasi pada data berukuran besar dan bertipe campuran yaitu data kategorik dan numerik. Metode *k-prototypes* ini menggabungkan ukuran jarak kedua metode non hierarki yaitu pada algoritma *k-means* dan algoritma *k-modes* (Ganmanah & Kudus, 2021).

Provinsi Jawa Barat merupakan provinsi kedua dengan kepadatan penduduk tertinggi menurut data BPS tahun 2021. Sekitar 5.312 desa tersebar di Jawa Barat dengan jumlah penduduk sekitar 50 juta diantaranya 39,8 juta jiwa tinggal di desa. Pembangunan desa merupakan kunci pertumbuhan ekonomi nasional seperti yang dipaparkan oleh Gubernur Jawa Barat, Ridwan Kamil (Antara, 2023). Oleh karena itu, diperlukan adanya informasi mengenai desa mana saja yang termasuk kategori tertinggal, berkembang, dan mandiri, agar pembangunan desa menjadi tepat sasaran.

Pada penelitian ini digunakan data campuran yaitu data dimensi Indeks Pembangunan Desa (IPD) pada data Potensi Desa (PODES) di Jawa Barat tahun 2021. IPD merupakan suatu ukuran yang disusun untuk menilai tingkat kemajuan atau perkembangan desa di Indonesia. IPD mencakup 5 Dimensi diantaranya : 1) Pelayanan Dasar, 2) Kondisi Infrastruktur, 3) Aksesibilitas/Transportasi, 4) Pelayanan Publik, dan 5) Penyelenggaraan Pemerintahan. Kelima dimensi tersebut tersusun dari 42 indikator, yang mana indikator-indikator ini akan mengklasifikasikan desa menjadi Desa Mandiri, Desa Berkembang, dan Desa Tertinggal (Harmadi, Suchaini, & Adji, 2020).

Beberapa penelitian sebelumnya telah dilakukan dalam analisis kluster menggunakan algoritma *k-prototypes*. Penelitian yang dilakukan oleh (Nooraeni., 2019) yaitu mengenai simulasi dengan menggunakan *k-prototypes*. Dari hasil penelitiannya disimpulkan bahwa secara umum algoritma *k-prototypes* dapat mempertahankan efisiensi algoritma *k-means* dalam menangani data berukuran besar tetapi menghilangkan keterbatasannya karena dapat digunakan pada data campuran. Sehingga *k-prototypes* memberikan hasil klustering yang lebih baik pada pembentukan kluster. Penelitian kedua dilakukan oleh (Wijayati & Saputro, 2023), membahas klustering data campuran numerik dan kategorik menggunakan algoritma *k-prototypes*, dimana disimpulkan bahwa *K-Prototypes* merupakan algoritma klustering untuk data bertipe campuran dengan ukuran kluster yang terbaik dapat dicari menggunakan metode *elbow*.

Penelitian yang dilakukan oleh (Fadilah & Wijayanto, 2023), membahas tentang perbandingan metode *One-Hot-Encoding*, *Gower Distance*, dan *K-Prototypes* berdasarkan akurasi pada data penyakit ginjal kronis, diperoleh bahwa *k-Prototypes* dengan $k = 2$ mampu menghasilkan *silhouette index* tertinggi dibandingkan empat metode lainnya, yaitu sebesar 0.3796. Penelitian selanjutnya oleh (Chrisinta, 2022) berjudul Identifikasi Karakteristik Desa di Provinsi Bengkulu Tahun 2018 Berdasarkan *Latent Class Cluster (LCC)*. Penelitian ini menggunakan peubah dari dimensi Indeks Pembangunan Desa (IPD) dan data dari Potensi Desa (PODES) tahun 2018. Kluster optimal yang diperoleh adalah 3 kluster, kluster terbaik ditunjukkan pada kluster 2 yang memberikan IPD yang relatif lebih tinggi daripada kluster 1 dan 3. Penelitian terakhir dilakukan oleh (Ganmanah & Kudus, 2021), membahas tentang penerapan algoritma *k-Prototypes* untuk mengelompokkan desa-desa di Jawa Barat berdasarkan Indeks Desa Membangun (IDM), pada penelitian ini diperoleh kesimpulan bahwa $k = 5$ merupakan jumlah kluster terbaik dengan nilai keagaman di dalam kluster terhadap keragaman antar kluster terkecil.

Penelitian ini menggunakan jumlah kluster berdasarkan klasifikasi IPD, berbeda dengan penentuan kluster pada penelitian – penelitian sebelumnya. Hal ini penting dilakukan agar kebijakan pemerintah efektif dan efisien pada desa – desa dengan dimensi IPD kategori tertinggal. Penelitian ini menerapkan

algoritma *K-Prototypes* yaitu untuk klasterisasi desa berdasarkan Indeks Pembangunan Desa. Data Indeks Pembangunan Desa yang digunakan merupakan data bertipe campuran (numerik dan kategorik). Algoritma *K-Prototypes* merupakan algoritma klastering untuk data bertipe campuran. Oleh karena itu, penelitian ini melakukan klasterisasi desa di provinsi Jawa Barat berdasarkan Indeks Pembangunan Desa (IPD) tahun 2021 menggunakan algoritma *K-Prototypes*. Tujuan penelitian ini yaitu sebagai masukan bagi pemerintah untuk mengeluarkan kebijakan yang tepat sasaran sesuai karakteristik dan potensi masing-masing desa menggunakan algoritma *K-Prototypes*.

METODOLOGI

Data

Data yang digunakan pada penelitian ini yaitu data PODES yang diperoleh dari Badan Pusat Statistik. Unit penelitian yang digunakan yaitu desa- desa di Provinsi Jawa Barat yang terdiri dari 4915 desa dari 19 kabupaten/kota. Penelitian ini menggunakan 57 peubah, peubah-peubah ini dipilih berdasarkan 5 dimensi IPD.

TABEL 1. Peubah yang digunakan.

Dimensi	Peubah	Keterangan
Pelayanan Dasar	X_1 : Jumlah sarana pendidikan	Numerik
	X_2 : Jumlah sarana kesehatan	Numerik
	X_3 : Jumlah tenaga kesehatan	Numerik
Kondisi Infrastruktur	X_4 : Jumlah bank pemerintah, swasta, dan pengkreditan	Numerik
	X_5 : Jumlah koperasi	Numerik
	X_6 : Jumlah kelompok pertokoan	Numerik
	X_7 : Jumlah pasar dengan Bangunan Permanen	Numerik
	X_8 : Jumlah pasar dengan Bangunan Semi Permanen	Numerik
	X_9 : Jumlah pasar tanpa bangunan	Numerik
	X_{10} : Jumlah minimarket/swalayan/supermarket	Numerik
	X_{11} : Jumlah restoran rumah makan	Numerik
	X_{12} : Jumlah warung Kedai makanan dan minuman	Numerik
	X_{13} : Jumlah hotel	Numerik
	X_{14} : Jumlah penginapan	Numerik
	X_{15} : Jumlah toko warung kelontong	Numerik
	X_{16} : Jumlah keluarga pengguna listrik PLN	Numerik
	X_{17} : Jumlah keluarga pengguna listrik Non-PLN	Numerik
	X_{18} : Jumlah keluarga bukan pengguna listrik	Numerik
	X_{19} : Penerangan di jalan utama desa yang menggunakan lampu tenaga surya	Kategorik (2 kategori)*
	X_{20} : Penerangan di jalan utama desa	Kategorik (3 kategori)*
	X_{21} : Bahan bakar untuk memasak yang digunakan oleh sebagianbesar keluarga	Kategorik (5 kategori)*
	X_{22} : Penggunaan fasilitas buang air besar sebagian besar keluarga di desa	Kategorik (4 kategori)*
	X_{23} : Sumber air untuk minum	Kategorik (10 kategori)*
	X_{24} : Sumber air untuk mandi/cuci	Kategorik (10 kategori)*
	X_{25} : Keberadaan warga yang menggunakan telepon seluler/handphone	Kategorik (3 kategori)*
	X_{26} : Fasilitas internet di kantor kepala desa	Kategorik (4 kategori)*

Dimensi	Peubah	Keterangan
Aksesibilitas/ Transportasi	X_{27} : Keberadaan kantor pos/pos pembantu/rumah pos	Kategorik (4 kategori)*
	X_{28} : Perusahaan agen jasa ekspedisi (pengiriman barang dokumen) swasta	Kategorik (4 kategori)*
	X_{29} : Lalu lintas ke desa	Kategorik (4 kategori)*
	X_{30} : Jenis permukaan jalan yang terluas	Kategorik (4 kategori)*
	X_{31} : Jalan darat dapat dilalui kendaraan bermotor roda 4 atau lebih	Kategorik (4 kategori)*
	X_{32} : Keberadaan angkutan umum	Kategorik (3 kategori)*
	X_{33} : Operasional angkutan umum yang utama	Kategorik (2 kategori)*
	X_{34} : Waktu tempuh dari kantor kepala desa ke kantor camat	Numerik
	X_{35} : Waktu tempuh dari kantor kepala desa ke kantor bupati/walikota	Numerik
	X_{36} : Biaya transportasi dari kantor kepala desa ke kantor camat (menit)	Numerik
Pelayanan Umum	X_{37} : Biaya transportasi dari kantor kepala desa ke kantor bupati/walikota	Numerik
	X_{38} : Ketersediaan fasilitas lapangan: sepak bola	Kategorik (4 kategori)*
	X_{39} : Ketersediaan fasilitas lapangan: bola voli	Kategorik (4 kategori)*
	X_{40} : Ketersediaan fasilitas lapangan: bulu tangkis	Kategorik (4 kategori)*
	X_{41} : Ketersediaan fasilitas lapangan: bola basket	Kategorik (4 kategori)*
	X_{42} : Ketersediaan fasilitas lapangan: tenis lapangan	Kategorik (4 kategori)*
	X_{43} : Ketersediaan fasilitas lapangan: tenis meja	Kategorik (4 kategori)*
	X_{44} : Ketersediaan fasilitas lapangan: futsal	Kategorik (4 kategori)*
	X_{45} : Ketersediaan fasilitas lapangan: renang	Kategorik (4 kategori)*
	X_{46} : Ketersediaan fasilitas lapangan: bela diri (pencak silat, karate, dll.)	Kategorik (4 kategori)*
Penyelenggaraan Pemerintahan	X_{47} : Ketersediaan fasilitas lapangan: bilyard	Kategorik (4 kategori)*
	X_{48} : Ketersediaan fasilitas lapangan: pusat kebugaran (fitnes, aerobik)	Kategorik (4 kategori)*
	X_{49} : Ketersediaan fasilitas lapangan: lainnya	Kategorik (4 kategori)*
	X_{50} : Jumlah aparatur pemerintahan: sekretariat desa	Numerik
	X_{51} : Jumlah aparatur pemerintahan: pelaksana teknis	Numerik
	X_{52} : Jumlah aparatur pemerintahan: pelaksana kewilayahan (kadus, rt, rw, dll.)	Numerik
	X_{53} : Jumlah aparatur pemerintahan: pegawai desa lainnya (hansip, dll)	Numerik
	X_{54} : Umur kepala desa	Numerik
	X_{55} : Umur sekretaris desa	Numerik
	X_{56} : Pendidikan tertinggi yang ditamatkan kepala desa	Kategorik (9 kategori)*
	X_{57} : Pendidikan tertinggi yang ditamatkan sekretaris desa	Kategorik (9 kategori)*

*masing-masing kategori dapat dilihat pada Buku IPD 2018 yang diterbitkan oleh BPS

Metode Penelitian

Analisis Kluster

Analisis kluster pertama kali dikembangkan oleh Tyron pada tahun 1939. Analisis kluster adalah metode populer untuk mengeksplorasi pola dalam populasi yang kompleks. Analisis kluster bertujuan

untuk mengelompokkan objek-objek ke dalam kluster yang memiliki kemiripan karakteristik sehingga objek dalam satu kluster menjadi sangat mirip, namun antar kluster akan berbeda. Analisis kluster merupakan teknik multivariat yang bertujuan untuk mengelompokkan objek-objek dengan karakteristik yang mirip ke dalam satu kluster dan objek dengan karakteristik yang berbeda ke kluster yang lain (Nuryani & Darwis, 2021).

Ukuran Kemiripan

Ukuran kemiripan yang digunakan dalam klustering adalah ukuran jarak (Wijayati & Saputro, 2023). Semakin besar jarak antara dua objek maka akan semakin berbeda atau tidak mirip kedua objek tersebut. Semakin kecil ukuran jarak, semakin kecil perbedaan antar objek satu dengan objek lainnya (Nooraeni, 2015). Ukuran jarak pada data campuran yang akan digunakan merupakan ukuran jarak *k-Prototypes*. Ukuran ini yaitu menggabungkan ukuran jarak pada data numerik dan ukuran data kategorik. Ukuran jarak algoritma *k-Prototypes* adalah sebagai (Annas, Irwan,S, dan Rais, 2022):

$$d_{ij} = \sum_{k=1}^p (x_{ik} + x_{jk})^2 + \gamma \sum_{k=p+1}^{p+m} \delta(x_{ik} + x_{jk}) \quad (1)$$

dimana:

d_{ij} : ukuran jarak antara objek i dan j

$\sum_{k=1}^p (x_{ik} + x_{jk})^2$: ukuran jarak untuk data numerik

γ : penyeimbang jarak

$\sum_{k=p+1}^{p+m} \delta(x_{ik} + x_{jk})$: ukuran jarak untuk data kategorik

K-Prototypes

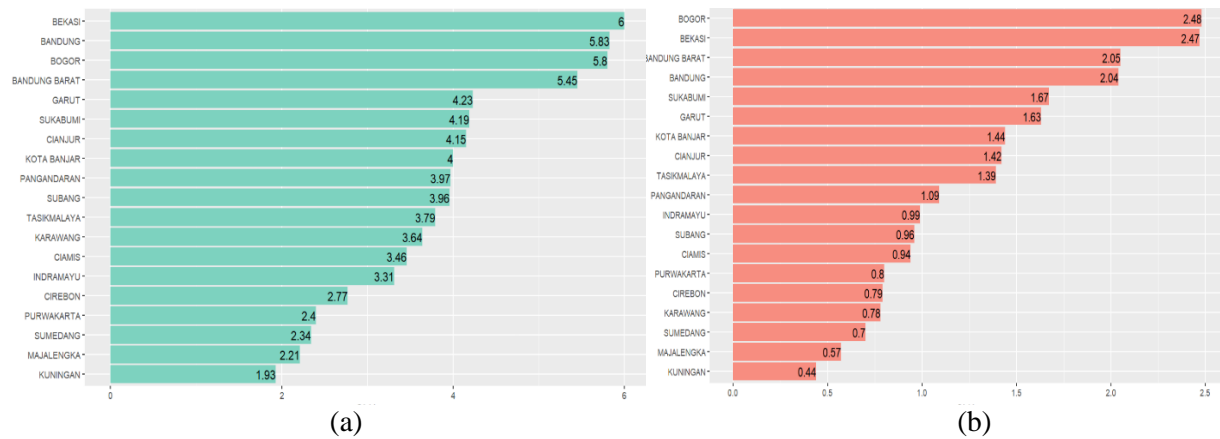
K-prototypes merupakan metode pengembangan antara *k-means* dan *k-modes* untuk mengatasi data dengan tipe campuran yaitu data numerik dan data kategorik (Iriawan, dkk., 2018). Pengembangan yang dilakukan oleh Huang untuk menjaga efisiensi algoritma *k-means* dalam menangani data berukuran besar dan dapat diterapkan pada numerik dan kategorik (Annas, dkk., 2022. *K-means* kluster merupakan metode kluster non hierarki yang bekerja dengan mempartisi objek kedalam satu atau lebih kluster, atau objek berdasarkan karakteristiknya (Ediyanto, Mara, & Satyahadewi, 2013). *K-modes* merupakan metode kluster non hierarki yang dapat digunakan pada data yang berbentuk kategorikal (nominal atau ordinal). Metode ini menggunakan ukuran modus untuk menentukan kluster sebagai centroid (Amah, Wahyuningsih, & Fidia, 2017). Jika *k-modes* menggunakan modus, perubahan centroid pada *k-means* menggunakan rata-rata. Perubahan yang mendasar terdapat pada ukuran kemiripan antara objek dengan centroid (*prototype*). Langkah-langkah yang dilakukan dalam melakukan klusterisasi menggunakan algoritma *k-prototype* yaitu sebagai berikut (Annas, Irwan, S, & Rais, 2022):

1. Menentukan banyaknya kluster (*k*) yang akan dibentuk.
2. Menentukan inisiasi awal *prototype*, dengan memilih sejumlah *k-prototype* secara acak dari *dataset*.
3. Melakukan perhitungan jarak semua observasi pada *dataset* terhadap *prototype* awal.
4. Mengalokasikan semua observasi ke dalam *dataset* ke dalam kluster yang memiliki jarak *prototype* terdekat dengan objek yang diukur.
5. Lakukan pengukuran ulang jarak antara semua observasi di dalam *dataset* terhadap *prototype* yang ada.
6. Merelokasikan semua observasi pada *dataset* terhadap *prototype* yang baru sampai titik pusat kluster konvergen.

HASIL DAN PEMBAHASAN

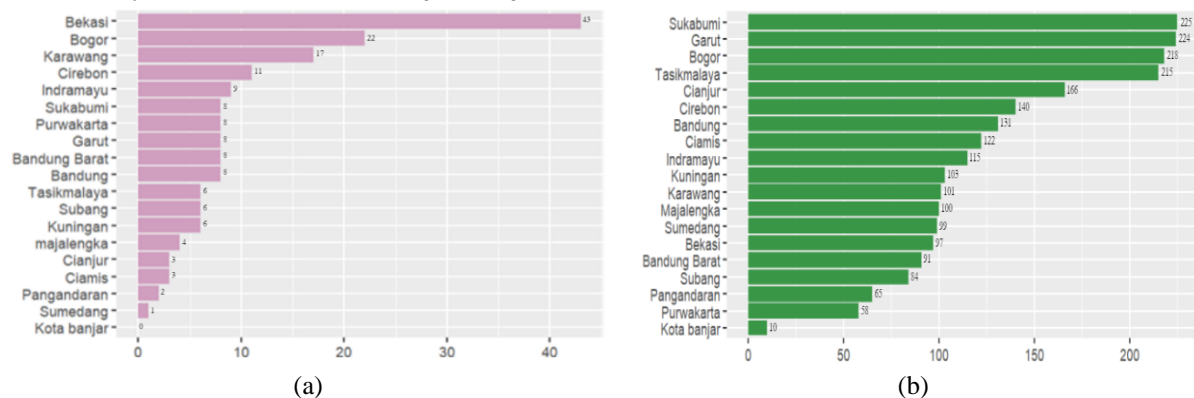
Eksplorasi Data

IPD pada dimensi ketersediaan pelayanan dasar mencakup kebutuhan dasar seperti pendidikan dan kesehatan. Sebaran sarana pendidikan yang diboboti oleh jumlah desa ditunjukkan oleh Gambar 1. Proporsi jumlah SMA per desa pada seluruh kota/kabupaten di Jawa Barat seperti yang ditunjukkan oleh Gambar 1.(a) tidak ada yang kurang dari satu, artinya jumlah SMA lebih banyak dibandingkan jumlah desa. Proporsi terkecil dimiliki oleh Kabupaten Kuningan yaitu 1,93 dan proporsi terbesar pada desa-



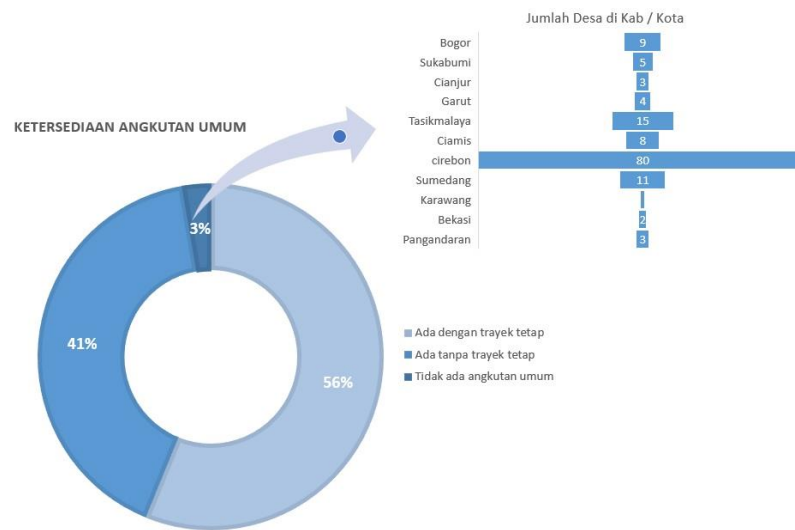
GAMBAR 1. (a) Proporsi Jumlah Sarana SMA per Jumlah Desa di Kabupaten/Kota, (b) Proporsi Jumlah Sarana SMP per Jumlah Desa di Kabupaten/Kota.

desa di Kabupaten Bekasi yaitu 6. Keterangan null maksudnya tidak ada pemerintahan administratif desa di kota tersebut. Namun berbeda dengan proporsi SMP pada Gambar 1.(b), terdapat daerah dengan proporsi yang kurang dari 0,5 yaitu Kabupaten Kuningan. Bila dilihat dari visualisasi peta, luas wilayah Kabupaten Kuningan tidak berbeda jauh dengan Kabupaten Bekasi. Menurut data dari BPS, rata-rata lama sekolah (RLS) Kabupaten Kuningan pada tahun 2021 yaitu 7.82 dan Kabupaten Bekasi 9.30. Artinya rata-rata penduduk Kuningan dalam mengenyam pendidikan tidak sampai tamat jenjang SMP. Kondisi ini perlu menjadi perhatian pemerintah dalam hal kebijakan untuk menaikkan angka RLS agar terlaksananya amanat ketuntasan wajib belajar 12 tahun.



GAMBAR 2. (a) Jumlah Sarana Rumah Sakit, (b) Jumlah Sarana Puskesmas.

Sebaran sarana kesehatan dapat dilihat pada Gambar 2. Banyak rumah sakit yang tidak berlokasi di desa, contohnya pada desa di Kota Banjar, Sumedang, dan Pangandaran. Namun sarana kesehatan lain seperti puskesmas sudah menyebar lokasinya di desa-desa seperti yang ditunjukkan oleh Gambar 2.(b).

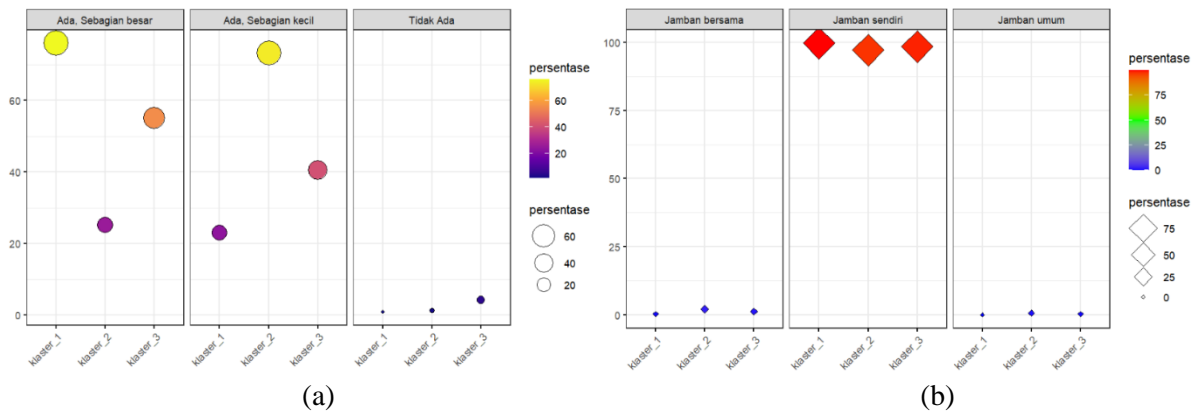


GAMBAR 3. Aksesibilitas Angkutan Umum.

Dimensi IPD juga memperhatikan peubah akses angkutan umum di desa. Aksesibilitas angkutan umum di desa cukup baik, hanya ada 3% desa seluruh Jawa Barat yang tidak memiliki akses angkutan umum seperti yang ditunjukkan oleh Gambar 3. Namun demikian cukup banyak desa-desa di Cirebon (80 desa) yang tidak memiliki akses angkutan umum, yaitu sekitar 20% dari total desa yang ada di Cirebon.

Klusterisasi

Klusterisasi dengan *k-prototypes* menghasilkan 931 desa yang termasuk kluster-1, 1880 desa kluster-2, dan 2104 desa kluster-3. Karakteristik desa pada setiap kluster menurut peubah kategorik dapat dilihat pada Gambar 4 dan Tabel 2. Desa-desa pada kluster-1 sebagian besar memiliki penerangan pada jalan-



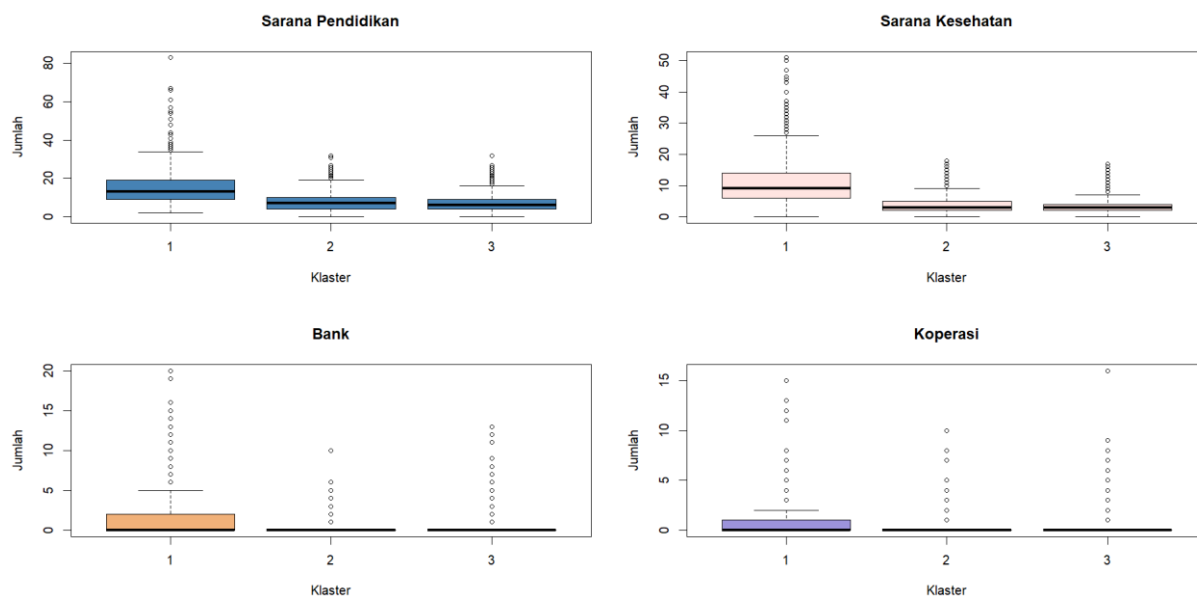
GAMBAR 4. (a) Persentase Keberadaan Penerangan di Jalan Utama Desa, (b) Persentase Fasilitas Jamban.

utamanya, dilanjutkan oleh kluster-3 kemudian kluster-2. Gambar 4.(a) pada kategori penerangan jalan ada, sebagian kecil, didominasi oleh kluster-2 dan persentase terkecil adalah kluster-1. Namun pada kategori tidak ada penerangan jalan, kluster-3 paling tinggi dan hampir tak ada jalan desa tanpa penerangan (0.8%) pada kluster-1. Fasilitas jamban di desa-desa pada kluster-1 sampai tiga di atas 97% berkategori jamban sendiri. Terdapat sekitar 2% desa pada kluster-2 dan 1% kluster-1 yang masih menggunakan jamban bersama. Sebesar 3.6% desa pada kluster-2 menggunakan kayu bakar sebagai bahan bakar memasak. Sebagian besar warga desa kluster-1 (99%) menggunakan gas elpiji bersubsidi.

TABEL 2. Karakteristik Klaster Berdasarkan Peubah Kategorik.

Peubah	Kategori	Persentase Klaster		
		1	2	3
X ₂₁ : Bahan Bakar Untuk Memasak yang Digunakan Oleh Sebagian besar Keluarga	2 : Elpiji 5,5 kg/blue gaz	0	0	0.04
	3 : Elpiji 12 kg	0.54	0.85	0.81
	4 : Elpiji 3 kg	99	96	96
	5 : Gas kota	0.1	0	0.1
	10 : Kayu Bakar	0.32	3.6	3.4
	1 : Air kemasan bermerek	0	0	0.01
	2 : Air isi ulang	50.2	42	18.2
	3 : Ledeng dengan meteran (PAM/PDAM)	5.1	4.4	5.3
	4 : Ledeng tanpa Meteran	0.3	0.9	1.1
	5 : Sumur bor atau pompa	30.2	26.5	20.5
X ₂₃ : Sumber Air Untuk Minum	6 : Sumur	7.1	8.03	33.3
	7 : Mata air	7.2	17.9	21.4
	8 : Sungai	0	0.32	0.14
	9 : Air hujan	0	0.05	0
	10 : Lainnya			
	X ₂₅ : Keberadaan Warga yang Menggunakan Telepon Seluler Handphone	1 : Sebagian besar warga	99.1	98.83
	2 : Sebagian kecil warga	0.86	1.1	1.38
	3 : Tidak ada	0.04	0.07	0.14
X ₂₆ : Fasilitas Internet Di Kantor Kepala Desa	1 : Berfungsi	97.1	89.2	90.26
	2 : Jarang berfungsi	0.64	1.9	1.57
	3 : Tidak berfungsi	0.54	1.7	1.47
	4 : Tidak ada	1.72	7.2	6.7
X ₃₀ : Jenis Permukaan Jalan yang Terluas	1 : Aspal/beton	99.7	97.2	97.4
	2 : Diperkeras	0.3	2.8	2.6
	3 : Tanah	0	0	0
	4 : Lainnya			
X ₃₈ : Ketersediaan Fasilitas Lapangan: Sepak Bola		59.7	24.15	57.03
		12.9	47.98	13.26
		2.58	5.32	5.13
		24.82	22.55	24.58
X ₄₁ : Ketersediaan Fasilitas Lapangan: Bola Basket	1 : Ada, baik	26.96	6.49	5.1
	2 : Ada, rusak sedang	3.54	2.23	1.33
	3 : Ada, rusak parah	0.97	0.53	0.57
	4 : Tidak ada	68.53	90.74	93.01
X ₄₅ : Ketersediaan Fasilitas: Renang		23.63	4.95	5.85
		1.07	0.48	0.57
		0.32	0.27	0.33
		74.97	94.31	93.23
X ₅₆ : Pendidikan tertinggi yang di tamatkan Kepala Desa	2 : Tidak tamat SD	0	0	0.05
	3 : Tamat SD	0	0.21	0
	4 : SMP	3.33	6.81	6.42
	5 : SMU	52.2	61.81	63.5
	6 :Akademi/DIII	3.97	4.36	3.42
	7 : Diploma IV/S1	36.63	24.2	24.9
	8 : S2	3.76	2.5	1.66
	9 :S3	0.1	0.1	0.05

Sumber air minum terbanyak untuk klaster-1 dan klaster-2 berasal dari air minum isi ulang, sedangkan bagi klaster-3 berasal dari air sumur. Sebanyak 98% warga desa menggunakan telepon seluler dan mayoritas fasilitas internet di kantor kepala desa berfungsi. Desa di klaster-2 memiliki kategori tidak ada fasilitas internet di kantor kepala desa paling banyak, yaitu 7.2%.



GAMBAR 5. Karakteristik Klaster Berdasarkan Peubah Numerik.

Karakteristik numerik setiap klaster dapat dilihat pada Gambar 5. Terlihat bahwa klaster-1 terdiri atas desa-desa dengan satatistik dari peubah numerik yang lebih tinggi dibandingkan dua klaster lainnya. Rata-rata jumlah sarana pendidikan desa-desa pada klaster-1 yaitu 15.2, klaster-2 dengan rata-rata 7.4, dan klaster-3 yaitu 7. Sarana kesehatan klaster-2 memiliki rata-rata satu satuan lebih tinggi dibanding klaster-3. Terdapat desa sebagai amatan ekstrim pada peubah koperasi, termasuk kedalam klaster-3. Desa tersebut memiliki jumlah koperasi satu satuan lebih banyak dibandingkan desa *outlier* atas anggota klaster-1. Karakteristik desa-desa klaster-3 ditandai dengan adanya jumlah keluarga bukan pengguna listrik di atas 22% yaitu desa Cikarang dan Campaka (Garut). Desa Campaka juga merupakan salah satu desa yang sebagian besar warganya menggunakan kayu bakar untuk memasak. Desa Campaka masih memanfaatkan air minum yang bersumber dari sumur yang dimasak dengan kayu bakar. Kayu bakar yang digunakan diperoleh dari kebun sehingga dapat lebih menghemat biaya.

KESIMPULAN DAN SARAN

Terdapat tiga klaster dalam pengelompokkan desa-desa di Jawa Barat yang terdiri atas 931 desa pada klaster-1, 1880 desa pada klaster-2, dan 2104 desa pada klaster-3. Karakteristik desa klaster-1 adalah desa-desa yang memiliki sarana kesehatan dan pendidikan yang jumlahnya memadai dan kondisi infrastruktur yang cukup baik. Klaster-2 memiliki rerata peubah numerik di bawah klaster-1 namun di atas klaster-3. Terdapat peubah kategorik dimana persentase klaster-3 sedikit lebih unggul dibandingkan klaster-2, yaitu pada peubah fasilitas internet di kantor kepala desa. Desa-desa dengan amatan ekstrim seperti pengguna non listrik lebih dari 20% yang termasuk anggota klaster-3 perlu peninjauan dan perhatian lebih lanjut untuk meningkatkan status desa menjadi desa mandiri.

REFERENSI

- Amah, N., Wahyuningsih, S., dan Fidia. (2017) 'Analisis Cluster Non-Hirarki dengan Menggunakan Metode K-Modes pada Mahasiswa Program Studi Statistika Angkatan 2015 FMIPA Universitas Mulawarman', *Jurnal Eksponensial*, Vol. 8(1), 9-15.
- Annas, S., Irwan, S, R. H., dan Rais, Z. (2022) 'K-Prototypes Algorithm For Clustering The Tectonic Earthquake In Sulawesi Island', *Jurnal Varian*, 5(2), 191-198.
- Annas, S., dkk. (2022) 'Using k-Means and Self Organizing Maps in Clustering Air Pollution Distribution in Makassar City', *Jambura Journal of Mathematics*, 4(1), 167-176.

- Chrisinta, D. (2022) 'Identifikasi Karakteristik Desa di Provinsi Bengkulu Tahun 2018 Berdasarkan Latent Class Cluster (LCC)', *Seminar Nasional Official Statistics, 1*, 927-936.
- Ediyanto, Mara, M. N., dan Satyahadewi, N. (2013) 'Pengklasifikasian Karakteristik dengan Metode K-Means Cluster Analysis', *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster), Vol. 02(2)*, 133-136.
- Fadilah, Z. R., dan Wijayanto, A. W. (2023) 'Perbandingan Metode Klasterisasi Data Bertipe Campuran: One-Hot-Encoding, Gower Distance, dan K-Prototype Berdasarkan Akurasi (Studi Kasus: Chronic Kidney Disease Dataset)', *Journal of Applied Informatics and Computing*, 63-73.
- Ganmanah, M., dan Kudus, A. (2021) 'Penerapan Algoritme K-Prototypes untuk Pengelompokan Desa-Desa di Provinsi Jawa Barat Berdasarkan Indikator Indeks Desa Membangun Tahun 2020', *Prosiding Statistika, Vol. 7(2)*.
- Harmadi, S. H., Suchaini, U., dan Adji, A. (2020) 'Indikator Pembangunan desa di Indonesia : Ditinjau dari Ketidaksiuaian Indikator Pembangunan Desa', Jakarta Pusat: Kertas Kerja.
- Iriawan, N., dkk. (2018) 'Ensemble Fuzzy, K-Prototypes Density Peaks Clustering Mixed) pada Pengelompokan Data Pelamar Bidikmisi Sejava-Timur Tahun 2016-2017', *Jurnal Ilmu Komputer dan Informasi, Vol. 11(2)*, 67.
- Nooraeni, R. (2015) 'Metode Cluster Menggunakan Kombinasi Algoritma Cluster K-Prototype dan Algoritma Genetika untuk data bertipe Campuran', *Jurnal Aplikasi Statistika Komputasi Statistik, Vol 7(2)*, 81-97.
- Nooraeni, R., dkk. (2019) 'K-Prototype untuk Pengelompokan Data Campuran', *Jurnal Statistika Universitas Muhammadiyah Semarang*, 1-6.
- Nuryani, I., Darwis, D. (2021) 'Analisis Clustering pada Pengguna Brand HP menggunakan Metode K-Means', *Prosiding Seminar Nasional Ilmu Komputer, Vol. 1(No. 1)*, 190-211.
- Trianasari, N., dkk. (2021), 'Application of beta mixture distribution in data on gpa proportion and course scores at the mbti telkom university', *Commun. Math Biol. Neurosci*, 1-12.
- Tyron, R.C. (1939), *Cluster Analysis*, Ann Arbor (MI): Edward Brothers.
- Wijayati, R., Saputro, D. R. (2023) 'Clustering Data Campuran Numerik dan Kategorik Menggunakan Algoritma K-Prototype', *PRISMA, Prosiding Seminar Nasional Matematika*, 702-706.
- Winata, Y. A. (2017) 'Penerapan Analisis Kovarians Multivariat pada bidang Gizi', *Jurnal Matematika, Vol. 6(No. 2)*, 1-10.